JOURNAL OF
COMPUTATIONAL AND
APPLIED MATHEMATICS

ELSEVIER Journal of Computational and Applied Mathematics 74 (1996) 231–244

Numerical methods for nonlinear inverse problems

T. Kärkkäinen, P. Neittaanmäki *, A. Niemistö

University of Jyväskylä, Department of Mathematics, Laboratory of Scientific Computing, FIN-40351 Jyväskylä, Finland

Received 1 July 1995

Abstract

Inverse problems of distributed parameter systems with applications to optimal control and identification are considered. Numerical methods and their numerical analysis for solving this kind of inverse problems are presented, main emphasis being on the estimates of the rate of convergence for various schemes. Finally, based on the given error estimates, a two-grid method and related algorithms are introduced, which can be used to solve nonlinear inverse problems effectively.

Keywords: Inverse problems; Finite element method; Error estimates; Two-grid method

1. Introduction and preliminaries

We consider inverse problems of distributed parameter systems, their numerical approximation with the finite element method, and, especially, introduce estimates of the rate of convergence for inverse identification and optimal control problems. In many applications, inverse problems can be nonlinear and ill-posed which makes them difficult to solve numerically. The parameter-to-observation mapping is often nonlinear and not invertible. Here, we study problems which can be stated in the following form:

$$\min_{q\in Q_{\mathrm{ad}},\ u\in U_{\mathrm{ad}}} F(u)+\beta G(q); \quad \text{subject to} \quad \mathbb{L}(q,u)=g \text{ in } H^{-1}, \tag{1}$$

where $F(u)$ and $G(q)$ are smooth convex functionals, β a given nonnegative real number, and Q_{ad}, U_{ad} sets for admissible controls and states, respectively.

There exist a lot of papers in the field of inverse problems concerning convergence studies, i.e., analyses of cases when the discretization parameter tends to zero ([1, 3, 4, 6, 14, 22, 24], and references therein). However, there are not so many papers containing error estimates [2, 9, 10, 12, 15, 17].

* Corresponding author.

PII S0377-0427(96)00026-X

Anyway, to develop effective numerical algorithms for inverse problems one needs to study the order of convergence. Error estimates in L^r and $W^{1,r}$-norms (and especially in L^∞ and $W^{1,\infty}$-norms) are important tools in studies of superconvergence, asymptotic error expansion and extrapolation techniques, in the analysis of a posteriori error estimates and grid refinement studies, etc. phenomena related with the design of modern numerical algorithms. Here we shall familiarize ourselves with the application of two-grid discretization of optimal control problems. For related works in the context of ordinary elliptic partial differential equations, see [27, 30–32].

The paper is organized as follows. In Section 2, we first give an abstract error estimate in the case $\beta = 0$. The method introduced uses a direct combination of the well-known output least-squares and equation error methods as a cost functional to be minimized. This abstract setting contains minimization problems for the second-order elliptic state equation with control and state constraints. The results are given under very loose restrictions for the operator, permitting nonlinearities and singularities, thus being applicable for the above-mentioned ill-posed problems. The abstract estimate developed is then applied to source identification problem with strongly nonlinear state equation and to the identification problem of a mildly nonlinear diffusion coefficient.

In Section 3, we restrict our study to unconstrained well-posed optimal control problems. Here, the cost functional will be coercive with respect to the unknown control because of the coercivity of $G(q)$ for $\beta > 0$. This theory allows wide variety of nonlinearities for the divergence form second-order state operator. These results are based on the first-order optimality system and its Taylor approximation. First, we give the L^r and $W^{1,r}$-error estimates which are then applied to the well-behaved counterpart of the above-mentioned source term identification problem. Due to the coercivity with respect to the control, we get one order better rate of convergence for the control as in the case of the identification problem. However, naturally parameter β is shown in this estimate.

As an application of the estimates developed, we give a two-grid approximation algorithms. By the given error estimates these approximations are demonstrated to be efficiently solvable.

Finally, we discuss about generalizations of these results to other problems, for instance, to the optimizing of the material parameter which can be considered as a well-behaved counterpart of the diffusion coefficient identification problem discussed in Section 2.

Standard notations for Sobolev spaces and associated norms are used. By H^{-1} we denote the dual space of H_0^1. Depending on the case, $\langle\cdot,\cdot\rangle$ denotes the L^2 inner product or the duality pairing between H^{-1} and H_0^1. By $|\cdot|$ we denote the Euclidian norm. Throughout this paper, C denotes a generic constant which may vary in different places, but is always independent of discretization parameters and β.

Let $\mathscr{T}_h$, $0 < h < 1$, be a family of triangulations of $\bar{\Omega}$. We assume that the family $\mathscr{T}_h$ is regular and quasi-uniform. For a nonnegative integer r we define a finite element space as

$$S_h^r = \left\{ v \mid v \in C^0(\bar{\Omega}),\ \ v|_T \in P_r\ \forall T \in \mathscr{T}_h \right\}, \tag{2}$$

where P_r is the space of less than or equal to r order polynomials. By $S_h^{r,0}$ we denote the subspace of S_h^r of functions which vanish on the boundary $\partial\Omega$. Approximation results with other useful properties of the finite element spaces S_h^r and a large amount of applications can be found in [7].

2. Error estimates for identification problems

Let us consider the least-squares formulated, constrained problem

$$\min_{q\in Q_{\mathrm{ad}},\ u\in U_{\mathrm{ad}}} \tfrac{1}{2}\|u-z\|^2_{H_0^1} \quad \text{subject to} \quad \mathbb{L}(q,u)=g \text{ in } H^{-1}. \tag{3}$$

Here z is an a priori observation of the true solution u, $Q_{\mathrm{ad}} \subset Q$ is a given set of admissible parameters, and $U_{\mathrm{ad}} \subset H_0^1$ the set of admissible states, respectively. $\mathbb{L} : Q \times H_0^1 \to H^{-1}$ is a general second-order nonlinear operator and Ω a bounded domain in $\mathbb{R}^2$ with smooth boundary $\partial\Omega$. In this section, we assume that problem (3) has at least one solution.

We assume that the operator $\mathbb{L}(\cdot,\cdot)$ is Lipschitz-continuous with respect to both arguments:

$$\forall \tilde{q} \in Q_{\mathrm{ad}} : \|\mathbb{L}(\tilde{q},v_1)-\mathbb{L}(\tilde{q},v_2)\|_{H^{-1}} \leqslant C\|v_1-v_2\|_{H_0^1} \quad \forall v_1,v_2 \in H_0^1, \tag{4}$$

$$\forall v \in U_{\mathrm{ad}} : \|\mathbb{L}(q_1,v)-\mathbb{L}(q_2,v)\|_{H^{-1}} \leqslant C\|q_1-q_2\|_Q \quad \forall q_1,q_2 \in Q. \tag{5}$$

Moreover, both assumptions (4) and (5) can be replaced with local Lipschitz-continuity (sometimes referred also as Lipschitz-continuity for bounded arguments) with respect to that argument for which $\mathbb{L}(\tilde{q},v)$ is coercive. Finally, we assume that between u and z we have an observation error of the form

$$\|u-z\|_{H_0^1} \leqslant \varepsilon. \tag{6}$$

Let $Q_{\mathrm{ad}}^{h_1}$ and $U_{\mathrm{ad}}^{h_2}$ be finite element approximation spaces of Q_{ad} and U_{ad} with discretization parameters h_1 and h_2, respectively. With these discrete spaces we introduce a cost functional

$$J(q_{h_1},u_{h_2}) = \tfrac{1}{2}\|u_{h_1}-z\|_{H_0^1} + \tfrac{1}{2}\|\mathbb{L}(q_{h_1},u_{h_2})-g\|^2_{H^{-1}}. \tag{7}$$

Cost functional (7) consists clearly of two parts: first part represents the usual least-squares formulation while the second part takes into account the original equation in a form that it is given. Hence, it combines directly the least squares and the so-called equation error formulations. The actual optimization problem can be now defined as

$$\begin{aligned} &\text{find } (q_{h_1},u_{h_2}) \in Q_{\mathrm{ad}}^{h_1} \times U_{\mathrm{ad}}^{h_2} \text{ such that } J(q_{h_1},u_{h_2}) \leqslant J(\tilde{q}_{h_1},\tilde{u}_{h_2}) \\ &\forall\ (\tilde{q}_{h_1},\tilde{u}_{h_2}) \in Q_{\mathrm{ad}}^{h_1} \times U_{\mathrm{ad}}^{h_2}. \end{aligned} \tag{8}$$

The following abstract estimate between the true pair (q,u) and the computed pair $(q_{h_1}^*,u_{h_2}^*)$ which is obtained as a solution of (8) holds.

Theorem 2.1. *Between (q,u) and $(q_{h_1}^*,u_{h_2}^*)$ we have an abstract estimate*

$$\begin{aligned} &\|u-u_{h_2}^*\|_{H_0^1} + \|\mathbb{L}(q,u)-\mathbb{L}(q_{h_1}^*,u_{h_2}^*)\|_{H^{-1}} \\ &\leqslant C\Big(\inf_{v_{h_2}\in U_{\mathrm{ad}}^{h_2}} \|u-v_{h_2}\|_{H_0^1} + \inf_{\bar{q}_{h_1}\in Q_{\mathrm{ad}}^{h_1}} \|q-\bar{q}_{h_1}\|_Q + \varepsilon\Big). \end{aligned} \tag{9}$$

Example 2.1 (*Identification of the source term*). For the first example, we use the above abstract estimate in the case of identification problem of source term q in the equation

$$\mathbb{L}(q,u) = -\nabla \cdot (\varphi(|\nabla u|^2)\nabla u) - q = g. \tag{10}$$

Here φ satisfies

$$0 < \alpha \leqslant \varphi(s) \quad \forall s \in \mathbb{R}^+, \ \varphi \text{ is locally Lipschitz}, \tag{11}$$

which guarantees that the whole operator $\mathbb{L}(\cdot,\cdot)$ has the required properties. For example, $\varphi = 1 + |\nabla u|^p$ for $p \geqslant 2$, fits into the framework of this and the next section. Let $U_{\text{ad}} = H_0^1$, $Q = H^{-1}$, and the set of admissible controls

$$Q_{\text{ad}} = \{\tilde{q} \in L^2 \mid \|\tilde{q}\|_{L^2} \leqslant \mu\}, \tag{12}$$

where μ is a given positive constant.

Corollary 2.1. *Assume that the true control q satisfies*

$$\|q\|_{L^2} \leqslant \mu - \delta \text{ for some } \delta > 0. \tag{13}$$

Let $u \in H^{r+1} \cap H_0^1$ and $q \in H^s$ in (10) *where $r \geqslant 1$, $s > 0$. Let $\bar{r} \geqslant r$ and $\bar{s} \geqslant \max\{s-1,0\}$ be positive integers, and choose $Q_{h_1} = S_{h_1}^{\bar{s}}$ and $U_{h_2} = S_{h_2}^{\bar{r},0}$. Then, between a computed control $q_{h_1}^*$ and q we have, for h_1 sufficiently small, an error estimate*

$$\|q - q_{h_1}^*\|_{H^{-1}} \leqslant C\left(h_1^{s+1} + h_2^r + \varepsilon\right). \tag{14}$$

Remark 2.1. Assumption (13) is a technical one guaranteeing that "for h_1 sufficiently small" there exists an element from $S_{h_1}^{\bar{s}}$ which is close to q (L^2 projection, for example) and belongs to Q_{ad}. For this purpose, we need q to be isolated from the constraint. Notice that this can be replaced with assuming directly that the L^2 projection of q belongs to Q_{ad} as in [3, Ch. VI.3].

Example 2.2 (*Identification of the nonlinear diffusion coefficient*). Second example where we apply the abstract estimate in Theorem 2.1, is an identification problem of a nonlinear parameter $q(u)$ in the quasi-linear equation

$$\mathbb{L}(q,u) = -\nabla \cdot (q(u)\,\nabla u) + b_0(\cdot, u, \nabla u) = g, \tag{15}$$

where b_0 contains the first-order terms.

Because q depends on u, we must introduce some kind of linearization. We make this as follows (see [15] for more details): assume that the observation $z \in C^0(\bar{\Omega}) \cap H_0^1$ (which is always the case in practice). Let us define an interval I as follows:

$$I = \left(\min_{x\in\Omega} z(x), \max_{x\in\Omega} z(x)\right). \tag{16}$$

Now we simply approximate the nonlinear parameter in a form $q_{h_1}(z)$ where $Q_{h_1}(I)$ is a finite element space on the one-dimensional interval I.

Let $U_{\mathrm{ad}} = H_0^1$, $Q = L^2$, and the set of discrete admissible parameters as

$$Q_{\mathrm{ad}}^{h_1} = \{\tilde{q}_{h_1}(z) \in Q_{h_1}(I) \mid \tilde{q}_{h_1}(z) \leqslant \lambda \text{ a.e. on } I\}, \tag{17}$$

where λ is a given positive constant. Notice that because we do not solve the actual equation (15) during the solution process, it is not required that q should be bounded from below.

Now we can state our second result in

Corollary 2.2. *Assume that the true parameter $q(\cdot)$ satisfies on I*

$$q \leqslant \lambda - \delta \quad \textit{for some } \delta > 0. \tag{18}$$

Let $u \in H_0^1 \cap H^{r+1} \cap W^{1,\infty}$ and $q(\cdot) \in W^{s,\infty}(\mathbb{R}) \cap C^{0,1}(\mathbb{R})$ in (15), where $r, s \geqslant 1$. Let $\bar{r} \geqslant r$, $\bar{s} \geqslant s-1$ be positive integers, and choose the discrete spaces $Q_{h_1}(I) = S_{h_1}^{\bar{s}}(I)$ and $U_{h_2} = S_{h_2}^{\bar{r},0}$. Then, between a computed parameter $q_{h_1}^*(z)$ and $q(u)$ we have, for h_1 sufficiently small, an error estimate

$$\|(q(u) - q_{h_1}^*(z))\, \nabla u\|_{L^2} \leqslant C\,(h_1^s + h_2^r + \varepsilon). \tag{19}$$

Remark 2.2. Notice that the estimate in Corollary 2.2 is of optimal order. Moreover, the weight ∇u in the above estimate is natural, because it indicates the well-known difficulties in the parameter identification problems in those parts of the domain Ω where ∇u vanishes.

The two examples above indicate clearly the nature of our abstract estimate in Theorem 2.1. Because the operator error there is only given with respect to the dual norm H^{-1}, the final estimate in different cases depends on whether we can replace this norm with some other (stronger) norm depending on $q, q_{h_1}^*$ and u as is the case with the second example but not with the first example. This weakness of the estimate is mainly caused by the lack of coercivity with respect to the unknown parameter. In the next section, we will find progress in the corresponding estimates for optimal control problems having this coercivity.

3. L^r- and $W^{1,r}$-error estimates for a class of Lagrange systems and its two-grid approximation

We study the saddle-point problem of the form

$$\begin{aligned}
&\mathbb{L}_{\mathrm{a}}(u) + (\mathbb{L}_{\mathrm{b}}'(u))^* p = f \quad \text{in } \Omega,\\
&\mathbb{L}_{\mathrm{b}}(u) - \mathbb{L}_{\mathrm{c}}(p) = g \quad \text{in } \Omega,\\
&u, p = 0 \quad \text{on } \partial\Omega.
\end{aligned} \tag{20}$$

Here $\mathbb{L}_{\mathrm{a}}$ and $\mathbb{L}_{\mathrm{b}}$ are second-order nonlinear divergence form operators and $\mathbb{L}_{\mathrm{c}}$ a nonlinear operator of order zero. By the prime, we denote the differentiation with respect to u. This type of saddle-point problems appear in the context of minimization problems with elliptic equation constraints. In the optimization terminology, it is usually called as a Lagrange system. To see the link to minimization problems, let us consider the following examples which are close to the examples studied in the previous section.

Example 3.1 (*Optimal control problem*).

$$\inf_{\substack{u\in H_0^1\\ q\in L^2}} \left\{ \tfrac{1}{2}\|u-z\|^2_{H_0^1} + \frac{\beta}{2}\|q\|^2_{L^2} \right\} \text{ subject to } \mathbb{L}_b(u) = Bq + g \text{ in } \Omega, \tag{21}$$

where the control operator B works from L^2 to L^2.

The first-order optimality conditions for (21) are following [20]:

$$\begin{aligned} -\Delta u + (\mathbb{L}_b'(u))^* p &= -\Delta z && \text{in } \Omega,\\ \mathbb{L}_b(u) - Bq &= g && \text{in } \Omega,\\ B^* p - \beta q &= 0 && \text{in } \Omega,\\ u, p &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{22}$$

After eliminating q, we have the system (20) with $\mathbb{L}_a(u) = -\Delta u$, $f = -\Delta z$, and $\mathbb{L}_c(p) = \frac{1}{\beta}BB^* p$.

There exists a large amount of problems where the system (20) represents one part of the optimality conditions [3, 6, 14, 20, 21]. Indeed, let us define the following problem of finding an optimal material parameter q.

Example 3.2 (*Optimal control of material parameter*).

$$\inf_{\substack{u\in H_0^1\\ q\in Q_{ad}}} \left\{ \frac{1}{2}\|u-z\|^2_{L^2} + \beta\left(\frac{1}{2+\lambda}\|\nabla q\|^{2+\lambda}_{L^{2+\lambda}} + \frac{1}{2}\|\nabla q\|^2_{L^2} \right) \right\}$$

subject to

$$\bar{\mathbb{L}}_b(q,u) = -\nabla\cdot(q\nabla u) + b_0(\cdot, u, \nabla u) = g, \tag{23}$$

where $Q_{ad} = \{\bar q \in W^{1,\infty} : \bar q \geqslant k > 0,\ \bar q_{|\partial\Omega} = q_{0|\partial\Omega}\}$ with some function $q_0 \in W^{2,r}$ and parameters $\beta > 0$ and $\lambda > 0$.

We associate the problem (23) with the following Lagrange functional:

$$\begin{aligned} \mathscr{L}(u,q,p) = {} & \tfrac{1}{2}\|u-z\|^2_{L^2} + \beta\left(\frac{1}{2+\lambda}\|\nabla q\|^{2+\lambda}_{L^{2+\lambda}} + \frac{1}{2}\|\nabla q\|^2_{L^2} \right) \\ & + \langle \bar{\mathbb{L}}_b(q,u) - g, p\rangle. \end{aligned} \tag{24}$$

If one assumes the problem having an isolated solution (u^*, q^*) in the interior of $H_0^1 \times Q_{ad}$, and z and g being sufficiently regular, then the first-order optimality conditions for (23) are

$$\begin{aligned} u^* + \bar{\mathbb{L}}_b'(u^*, q^*, p^*) &= z && \text{in } \Omega,\\ \bar{\mathbb{L}}_b(q^*, u^*) &= g && \text{in } \Omega,\\ \nabla u^* \cdot \nabla p^* + \beta(-\nabla\cdot((1+|\nabla q^*|^\lambda)\nabla q^*)) &= 0 && \text{in } \Omega,\\ u^*, p^* = 0,\ q^* &= q_0 && \text{on } \partial\Omega. \end{aligned} \tag{25}$$

The first two equations of this optimality system fits into the framework of system (20) with fixed q and $\mathbb{L}_a(u) = u$, $\mathbb{L}_b(u) = \bar{\mathbb{L}}_b(q,u)$, $f = z$, and $\mathbb{L}_c(p) = 0$.

The results contain also the augmented Lagrange formulated problems when we have, for instance, instead of (24) the Lagrangian

$$\bar{\mathscr{L}}(u,q,p) = \mathscr{L}(u,q,p) + \frac{\gamma}{2}\|\bar{\mathbb{L}}_{\mathbf{b}}(q,u) - g\|^2_{H^{-1}}. \tag{26}$$

This turns out the optimality system where one has, instead of $\mathbb{L}_{\mathbf{a}}(u) = u$, the operator

$$\mathbb{L}_{\mathbf{a}}(u) = u + \gamma(\bar{\mathbb{L}}'_{\mathbf{b}}(q^*,u^*))^*[(-\Delta)^{-1}(\bar{\mathbb{L}}_{\mathbf{b}}(q,u) - g)]. \tag{27}$$

The advantage of the formulation (27) compared with (24) is the fact that the optimality system will be strongly monotone with respect to the variable u at the solution u^* whenever the penalty parameter γ is large enough.

Note that a suitable regularization of the identification problems studied in the previous section leads to the similar formulations as in Examples 3.1 and 3.2. Numerical studies related with these problems can be found in [3,9,12,14,24], for example. The estimates for the discretization error attained in these papers have been carried out in Hilbert norms and no results can be found for the whole L^r- and $W^{1,r}$-scale. Although we perform our studies without any additional constraints, there exists still plenty of substantial applications.

Next we present the error estimates in L^r- and $W^{1,r}$-norms for the continuous piecewise linear finite element discretization of the system (20) in bounded, convex polygonal domains in $\mathbb{R}^2$. We mention that the results can be extended to higher-order finite element spaces [11,29] and to domains with smooth boundary [26,28,29].

3.1. L^r- and $W^{1,r}$-error estimates

Let $\mathbb{L}_{\mathbf{a}}(u)$, $\mathbb{L}_{\mathbf{b}}(u)$, $\mathbb{L}_{\mathbf{c}}(p) : H_0^1 \to H^{-1}$ be nonlinear operators of the form

$$\begin{aligned}
\mathbb{L}_{\mathbf{a}}(u) &= -\nabla\cdot(\mathbf{A}_1(x,u,\nabla u)) + \mathbf{A}_0(x,u,\nabla u) && \text{in } \Omega,\\
\mathbb{L}_{\mathbf{b}}(u) &= -\nabla\cdot(\mathbf{B}_1(x,u,\nabla u)) + \mathbf{B}_0(x,u,\nabla u) && \text{in } \Omega,\\
\mathbb{L}_{\mathbf{c}}(p) &= \mathbf{C}_0(x,p) && \text{in } \Omega.
\end{aligned} \tag{28}$$

Let $B^{\hat{R}}(\hat{u})$ denote the following closed convex subset of H_0^1 :

$$B^{\hat{R}}(\hat{u}) = \{w \in H_0^1 \cap W^{1,\infty}:\ \|\hat{u} - w\|_{W^{1,\infty}} \leqslant \hat{R}\}. \tag{29}$$

Let $\hat{u}$, $\hat{p} \in H_0^1$ and $\hat{R}$ be fixed. We make the basic assumptions

Assumption 3.1. (1) $\mathbf{A}_1(x,z,\xi) : \bar{\Omega}\times\mathbb{R}^1\times\mathbb{R}^2 \mapsto \mathbb{R}^2$ and $\mathbf{A}_0(x,z,\xi) : \bar{\Omega}\times\mathbb{R}^1\times\mathbb{R}^2 \mapsto \mathbb{R}^1$ are $C^2(\mathbb{R}^1\times\mathbb{R}^2)$ such that the second derivatives are locally Lipschitz functions with respect to the variables z and ξ at every $x \in \Omega$ and $\partial_\xi \mathbf{A}_1(x,w(x),\ \nabla w(x)) \in W^{1,\infty}$ and $\partial_z \mathbf{A}_1(x,w(x),\ \nabla w(x))$, $\partial_z \mathbf{A}_0(x,w(x),\nabla w(x))$, $\partial_\xi \mathbf{A}_0(x,w(x),\ \nabla w(x))$ belongs to L^∞ whenever $w \in W^{1,\infty}$.

(2) $\mathbf{B}_1(x,z,\xi) : \bar{\Omega} \times \mathbb{R}^1 \times \mathbb{R}^2 \mapsto \mathbb{R}^2$ and $\mathbf{B}_0(x,z,\xi) : \bar{\Omega} \times \mathbb{R}^1 \times \mathbb{R}^2 \mapsto \mathbb{R}^1$ are $C^3(\mathbb{R}^1 \times \mathbb{R}^2)$ such that the third derivatives are locally Lipschitz functions in z and ξ, and $\mathbf{B}_1$ continuous and $\mathbf{B}_0$ bounded in x for fixed z, ξ. Moreover,

$$\langle \mathbb{L}'_{\mathbf{b}}(w)v, v\rangle \geqslant C\ \|v\|^2 \quad \forall v \in H_0^1,\ w \in B^{\hat{R}}(\hat{u}), \tag{30}$$

and, particularly,

$$\eta^{\top}(\partial_\xi \mathbf{B}_1(x,w,\nabla w))\eta \geqslant C\ |\eta|^2 \quad \forall \eta \in \mathbb{R}^2,\ w \in B^{\hat{R}}(\hat{u}). \tag{31}$$

(3) $\mathbb{L}_{\mathbf{a}}'(w) + ((\mathbb{L}_{\mathbf{b}}'(w))^* p)'$ is positive semidefinite for all $w \in B^{\hat{R}}(\hat{u})$.

(4) $\mathbf{C}_0(x,z)\ :\ \Omega \times \mathbb{R}^1 \mapsto \mathbb{R}^1$ a two times differentiable function with respect to z such that $0 \leqslant \partial_z \mathbf{C}_0(x,w(x)) \in L^\infty$ for all $w \in B^{\hat{R}}(\hat{p})$.

(5) There exists a locally unique solution (u^*,p^*) in $B^{R_1}(\hat{u}) \cap W^{2,r} \times B^{R_1}(\hat{p}) \cap W^{2,r}$, $R_1 < \hat{R}$, for the problem (20).

It is worth noticing that these assumptions have a local nature.

Now, let us turn to the question of the piecewise linear finite element approximation of the problem (20). We define the finite element space $V_h \stackrel{\text{def}}{=} S_h^{1,0}$ and further the FE-approximation of (20) by

$$\begin{aligned} &\langle \mathbb{L}_{\mathbf{a}}(u_h), v_h\rangle + \langle (\mathbb{L}_{\mathbf{b}}'(u_h))^* p_h, v_h\rangle = \langle f, v_h\rangle \quad \forall v_h \in V_h, \\ &\langle \mathbb{L}_{\mathbf{b}}(u_h), w_h\rangle - \langle \mathbb{L}_{\mathbf{c}}(p_h), w_h\rangle = \langle g, w_h\rangle \quad \forall w_h \in V_h. \end{aligned} \tag{32}$$

Let R be a sufficiently small real number, such that $B^R(u^*) \subset B^{\hat{R}}(\hat{u})$ and $B^R(p^*) \subset B^{\hat{R}}(\hat{p})$. Then we have the following approximation result.

Theorem 3.1. *Let $(u^*,p^*) \in (W^{2,r} \cap H_0^1)^2$, $r \geqslant 2+\varepsilon$, $\varepsilon > 0$, be the solution of the problem* (20). *Then, for h sufficiently small, the problem* (32) *has a solution $(u_h^*, p_h^*) \in V_h \cap B^R(u^*) \times V_h \cap B^R(p^*)$ satisfying the sharp estimates*

$$\begin{aligned} &\|u^* - u_h^*\|_{W^{1,r}} + \|p^* - p_h^*\|_{W^{1,r}} \leqslant C\ h, \quad r \in [2+\varepsilon, \infty], \\ &\|u^* - u_h^*\|_{L^r} + \|p^* - p_h^*\|_{L^r} \leqslant C\ h^2, \quad r \in [2+\varepsilon, \infty), \\ &\|u^* - u_h^*\|_{L^\infty} + \|p^* - p_h^*\|_{L^\infty} \leqslant C\ h^2\, |\log h|. \end{aligned} \tag{33}$$

3.2. Applications to optimal control problems

With Example 3.1 in mind, we consider the optimal control problems governed by strongly nonlinear state equations. We recall the operators of the formula (22) and give the assumptions under which Assumption 3.1 is fulfilled. First of all, we give the following definitions and assumptions:

$$\begin{aligned} &\mathbb{L}_{\mathbf{a}}(u) \stackrel{\text{def}}{=} -\Delta u,\ \mathbb{L}_{\mathbf{b}}(u) \stackrel{\text{def}}{=} -\nabla \cdot (\varphi(|\nabla u|^2)\nabla u), \\ &\varphi \in C^3,\ \varphi(s) \geqslant \alpha > 0,\ \frac{\partial \varphi}{\partial s} \geqslant 0,\ \forall s \in \mathbb{R}^+,\ \frac{\partial^2 \varphi}{\partial s^2} \text{ locally Lipschitz,} \end{aligned} \tag{34}$$

$$\begin{aligned} &\mathbb{L}_{\mathbf{c}}(p) \stackrel{\text{def}}{=} \frac{1}{\beta} B B^* p, \\ &Bq \in C^{1,\gamma}(\bar{\Omega}) \quad \forall q \in L^2,\ B^* p \in W^{2,r} \quad \forall p \in W^{2,r}, \\ &\langle Bs, s\rangle \geqslant 0,\ \langle Bs, v\rangle \leqslant C\ \|s\|_{H^1} \|v\|_{H^1}, \quad \forall s, v \in H_0^1, \end{aligned} \tag{35}$$

and

$$f \stackrel{\text{def}}{=} -\Delta z \in L^r, g \in C^{1,\gamma}(\bar{\Omega}). \tag{36}$$

Here β, γ, and α are strictly positive real numbers. The parameter $r > 2$ will be fixed later.

Under these assumptions problem (21) has a solution $(u^*, q^*) \in H_0^1 \times L^2$. This one can infer by standard functional analytical convergence studies (see [5, 20], for example). Also the existence of a Lagrange variable $p^* \in H_0^1$ satisfying the optimality system (22) is guaranteed.

Let us recall the optimality system in the following form:

$$\begin{aligned}
&-\Delta u - \nabla\cdot\left(\left[\varphi(|\nabla u|^2) + 2\frac{\partial\varphi}{\partial s}(|\nabla u|^2)(\nabla u)^2\right]\nabla p\right) = -\Delta z \quad \text{in } \Omega,\\
&-\nabla\cdot(\varphi(|\nabla u|^2)\nabla u) - \frac{1}{\beta}BB^* p = g \quad \text{in } \Omega,\\
&B^* p - \beta q = 0 \quad \text{in } \Omega,\\
&u, p = 0 \quad \text{on } \partial\Omega,
\end{aligned} \tag{37}$$

where $(\nabla u)^2 = \nabla u \otimes \nabla u \in \mathbb{R}^{2\times 2}$ with $\otimes$ denoting the tensor product of two vectors in R^2. The first two equations of this system are in the category of the saddle-point problems discussed in the previous subsection.

The existence of the solution (u^*, p^*) for the problem

$$\text{find } (u, p) \in H_0^1 \times H_0^1 \text{ satisfying (37)} \tag{38}$$

follows from the above discussion.

Let q_h^* be the solution of the problem

$$\text{find } q_h^* \in V_h \text{ such that } \left\langle q_h - \frac{1}{\beta}B^*(p_h^*), v_h\right\rangle \quad \forall v_h \in V_h. \tag{39}$$

For the FE approximation of (38) and (39) we have the following theorem.

Theorem 3.2. *Let assumptions* (34)–(36) *hold and* $\|\Delta(u^* - z)\|_{L^r}$ *be small enough. Then problem* (21) *has a solution* (u^*, q^*) *in* $(H_0^1 \cap W^{2,r}) \times W^{2,r}$ *with* $2 < r < 2/(2 - \pi/\omega)$, *and* ω *the measure of the largest inner angle of* Ω. *Moreover, there exists a Lagrange variable* p^* *in* $H_0^1 \cap W^{2,r}$ *associated with the state equation constraint, such that* (u^*, q^*, p^*) *satisfies* (37) *and* (u^*, p^*) *is an isolated solution of* (38). *Further, for h sufficiently small, there exists a solution* (u_h^*, p_h^*) *for the piecewise linear FE approximation of problem* (38) *and* q_h^* *of problem* (39) *satisfying the asymptotically optimal error estimates* (33) *and*

$$\begin{aligned}
&\beta\|q^* - q_h^*\|_{L^r} \leqslant C\, h^2, \qquad r \in [2, \infty),\\
&\beta\|q^* - q_h^*\|_{L^\infty} \leqslant C\, h^2 |\log h|, \quad r = \infty.
\end{aligned} \tag{40}$$

Remark 3.1. Analogously to Theorem 3.1, Theorem 3.2 holds also if the properties (34) holds only locally at u^*.

Remark 3.2. The assumption $\|\Delta(u^* - z)\|_{L^r}$ being small, let say less than or equal to $\eta = \eta(\nabla u^*)$ means, roughly speaking, an attainability condition of the η neighborhood of z in $W^{2,r}$-norm. In context of identification problems this assumption is usually argued to hold for sufficiently small regularization parameter β.

Remark 3.3. Although the optimal control q^* was originally hunted in L^2, we got its $W^{2,r}$ regularity. This gave us the reason to use the piecewise linear approximation space for the control.

Remark 3.4. The regularity assumptions for B, $Bq \in C^{1,\gamma}(\bar{\Omega})$ $\forall q \in L^2$ and $B^* p \in W^{2,r}$ $\forall p \in W^{2,r}$ are fulfilled, for example, if B is defined by the convolution with kernel in $W^{2,r}$.

Since the assumption $Bq \in C^{1,\gamma}(\bar{\Omega})$ $\forall q \in L^2$ was needed only to achieve the $W^{2,r}$-regularity for u^*, this assumption as well as the assumption for g becomes weaker, if the character of the nonlinearity of the state equation is milder, or the boundary of the domain is smooth. If, for instance, $\partial\Omega$ is smooth, it is enough that $Bq \in C^{0,\gamma}(\bar{\Omega})$ $\forall q \in L^2$ [13]. If, in addition, the principal part of the operator $\mathbb{L}_{\mathbf{b}}(u)$ is mildly nonlinear, $Bq \in L^r$ $\forall q \in L^r$ is enough [18]. The same holds if the operator $\mathbb{L}_{\mathbf{b}}(u)$ is defined by a C^1 strongly coercive vector field (see [16], Ch. IV, Definition 3.1). This condition is being fulfilled by including an additional assumption

$$\varphi(s) + 2\frac{\partial\varphi(s)}{\partial s} \cdot s \leqslant M < \infty \ \forall s \in \mathbb{R}^+$$

in (34).

Remark 3.5. In assumption (36), it is enough to assume $f \in H^{-1}$ such that $\langle f, v\rangle = \langle \nabla z, \nabla v\rangle$ for every $v \in H_0^1$ and $z \in H_0^1 \cap W^{1,\infty}$ (See [27, Remark 2.1]).

3.3. Two-grid approximation of optimal control problems

In this section, we shall see that the complexity of solving the nonlinear problem (21) can be reduced approximately to the complexity of solving one or two linear systems at most the same size. This we will attain by two-grid algorithms. We give algorithms presented for mildly nonlinear elliptic problems in [32].

For any H_0^1 function u, p, v, w, $\bar{u}$, and $\bar{p}$ we define the following forms:

$$\mathbb{A}(u,v) \stackrel{\text{def}}{=} \langle \mathbb{L}_{\mathbf{a}}(u), v\rangle = \langle \nabla u, \nabla v\rangle,$$

$$\mathbb{B}(u,w) \stackrel{\text{def}}{=} \langle \mathbb{L}_{\mathbf{b}}(u), w\rangle = \langle \varphi(|\nabla u|^2)\nabla u, \nabla w\rangle,$$

$$\mathbb{B}_{\bar{u}}(p,v) \stackrel{\text{def}}{=} \langle (\mathbb{L}_{\mathbf{b}}'(\bar{u}))^* p, v\rangle = \left\langle \left[\varphi(|\nabla \bar{u}|^2) + 2\frac{\partial\varphi}{\partial s}(|\nabla \bar{u}|^2)(\nabla \bar{u})^2\right] \nabla p, \nabla v\right\rangle,$$

$$\mathbb{B}_{\bar{u}\bar{p}}(u,v) \stackrel{\text{def}}{=} \langle ((\mathbb{L}_b'(\bar{u}^*))^* p^*)' u, v\rangle = 2\left\langle \left[\left(2\frac{\partial^2\varphi}{\partial s^2}(|\nabla\bar{u}|^2)(\nabla\bar{u})^2 \right.\right.\right.$$

$$\left.\left.\left. + \frac{\partial\varphi}{\partial s}(|\nabla\bar{u}|^2)\right)\nabla\bar{u}\cdot\nabla\bar{p} + 2\frac{\partial\varphi}{\partial s}(|\nabla\bar{u}|^2)\nabla\bar{u}\otimes\nabla\bar{p}\right]\nabla u, \nabla v\right\rangle,$$

$$\mathbb{C}(p,w) \stackrel{\text{def}}{=} \langle \mathbb{L}_c(p), w\rangle = \frac{1}{\beta}\langle BB^* p, w\rangle.$$

Let h be the mesh parameter as in the previous section and let $H > h$ be another mesh size parameter defining the finite element space V_H. We define the following two-grid algorithm.

Algorithm 3.1. (1) Find $(u_H, p_H) \in V_H \times V_H$ such that

$$\mathbb{A}(u_H, v_H) + \mathbb{B}_{u_H}(p_H, v_H) = \langle -\Delta z, v_H\rangle \quad \forall v_H \in V_H,$$
$$\mathbb{B}(u_H, w_H) - \mathbb{C}(p_H, w_H) = \langle g, w_H\rangle \quad \forall w_H \in V_H.$$

(2) Find $(u_h, p_h) \in V_h \times V_h$ such that $\forall v_h, w_h \in V_h$

$$\mathbb{A}(u_h, v_h) + \mathbb{B}_{u_H p_H}(u_h - u_H, v_h) + \mathbb{B}_{u_H}(p_h, v_h) = -\langle \Delta z, v_h\rangle,$$
$$\mathbb{B}_{u_H}(u_h - u_H, w_h) - \mathbb{C}(p_h, w_h) = -\mathbb{B}(u_H, w_h) + \langle g, w_h\rangle.$$

(3) Find $q_h \in V_h$ such that

$$\left\langle q_h - \frac{1}{\beta}B^*(p_h^*), s_h\right\rangle \quad \forall s_h \in V_h.$$

By Taylor expansion formula it can be seen that for the solution (u_h, p_h) of this algorithm the following estimate is valid:

$$\|u^* - u_h\|_{H^1} + \|p^* - p_h\|_{H^1} \leqslant C\Big(\inf_{v_h\in V_h}\|u^* - v_h\|_{H^1} + \inf_{w_h\in V_h}\|p^* - w_h\|_{H^1}$$
$$+\|u^* - u_H\|^2_{H^{1,4}} + \|p^* - p_H\|^2_{H^{1,4}}\Big).$$

Hence, by Theorem 3.2 and the stability of L^2 projections in H^1-norm [8], we have the following theorem.

Theorem 3.3. *Let $(u^*, p^*, q^*) \in (W^{2,4})^3$ be the solution of the system* (37). *Then, for h sufficiently small, the solution (u_h, p_h, q_h) of Algorithm* 3.1 *satisfies the following estimate*:

$$\|u^* - u_h\|_{H^1} + \|p^* - p_h\|_{H^1} + \beta\|q^* - q_h\|_{H^1} \leqslant C\left(h + H^2\right). \tag{41}$$

Hence, taking $H = \sqrt{h}$ Algorithm 3.1 gives an approximate solution of the optimality system (37) satisfying the optimal error estimate

$$\|u^* - u_h\|_{H^1} + \|p^* - p_h\|_{H^1} + \beta\|q^* - q_h\|_{H^1} \leqslant C\,h. \tag{42}$$

Unfortunately, Algorithm 3.1 does not give progress enough to get the optimal order of convergence in L^r-norms. However, this will be reached after taking two Newton steps on the fine-grid, i.e., by replacing (2) in Algorithm 3.1 with

(2′) Find $(\bar{u}_h, \bar{p}_h) \in V_h \times V_h$ such that $\forall v_h, w_h \in V_h$

$$\mathbb{A}(\bar{u}_h, v_h) + \mathbb{B}_{u_H p_H}(\bar{u}_h - u_H, v_h) + \mathbb{B}_{u_H}(\bar{p}_h, v_h) = -\langle \Delta z, v_h \rangle,$$
$$\mathbb{B}_{u_H}(\bar{u}_h - u_H, w_h) - \mathbb{C}(\bar{p}_h, w_h) = -\mathbb{B}(u_H, w_h) + \langle g, w_h \rangle.$$

Find $(u_h, p_h) \in V_h \times V_h$ such that $\forall v_h, w_h \in V_h$

$$\mathbb{A}(u_h, v_h) + \mathbb{B}_{\bar{u}_h \bar{p}_h}(u_h - \bar{u}_h, v_h) + \mathbb{B}_{\bar{u}_h}(p_h, v_h) = -\langle \Delta z, v_h \rangle,$$
$$\mathbb{B}_{\bar{u}_h}(u_h - \bar{u}_h, w_h) - \mathbb{C}(p_h, w_h) = -\mathbb{B}(\bar{u}_h, w_h) + \langle g, w_h \rangle.$$

By similar arguments as used for Algorithm 3.1 we have:

Theorem 3.4. *Let $(u^*, p^*, q^*) \in (W^{2,\infty})^3$ be the solution of the system* (37). *Then, for h sufficiently small, the solution (u_h, p_h, q_h) of Algorithm* 3.1 *with* (2′) *satisfies the following estimate*:

$$\|u^* - u_h\|_{L^\infty} + \|p^* - p_h\|_{L^\infty} + \beta\|q^* - q_h\|_{L^\infty}$$
$$\leqslant C\left(|\log h| h^2 + |\log h|^2 H^4\right). \tag{43}$$

The previous theorem tells us that taking $H = \sqrt{h}$ the asymptotic rate of the convergence in L^r-norms using Algorithm 3.1 with (2′) to approximate the optimality system is $O(h^2|\log h|^2)$. Hence, we get a nearly optimal convergence rate by solving one nonlinear problem of size $\frac{2^2}{h}$, two linear problems of size $\frac{2^2}{h^2}$, and one linear problem of size $\frac{1}{h^2}$. For instance, if $H = \frac{1}{16}$, then $h = \frac{1}{256}$ and one must solve one nonlinear problem with 900 unknowns two linear problems with 260 100 unknowns and another linear problem with 65 025 unknowns instead of solving the nonlinear problem with 585 225 unknowns.

3.4. Extensions

Similar results as given in the previous sections hold also for other formulations of (21) and for other problems similar to (21). As, for example, for the saddle-point problem associated with the augmented Lagrange functional corresponding to the problem (21) or the problem of Example 3.2. In the latter case, for $(u^*, p^*, q^*) \in W^{2,r}$, the following estimates were shown in [26]:

$$\|u^* - u_h\|_{W^{1,r}} + \|p^* - p_h\|_{W^{1,r}} + \beta\|q^* - q_h\|_{W^{1,r}} \leqslant C\left(h + \frac{h^{2+\frac{4}{r}}}{\beta^2}\right),$$
$$\|u^* - u_h\|_{L^r} + \|p^* - p_h\|_{L^r} + \beta\|q^* - q_h\|_{L^r} \leqslant C\left(h^{\min\{2,3-\frac{4}{r}\}} + \frac{h^{2+\frac{4}{r}}}{\beta^2}\right). \tag{44}$$

The latter estimate is not sharp. Sharp result, similar to that of Theorem 3.1, can be proved using a more careful analysis as in [25].

The optimal choice of the grid-sizes in two-grid algorithms may differ from that presented in the previous section. For example, if the state equation in Example 3.1 would be mildly nonlinear, such that the nonlinearity would depend only on u and not on its gradient, the algorithm corresponding to Algorithm 3.1 would give an approximate solution with the order in H^1-norm same as that of the discretization error by using the grid-sizes $H = h^{1/3}$. The nearly optimal order of approximation error in L^r-norms will be reached by the same algorithm using the grid-sizes $H = h^{1/2}$, whereas for strongly nonlinear problem this is generally not attainable.

In cases where the control cannot be eliminated from the first two equations of the optimality conditions in (22), the analysis must be performed for the whole optimality system. This will effect also to the two-grid discretization. If, for instance, the third equation in the optimality system (22) would be mildly nonlinear, we could use a two-grid algorithm associated with Algorithm 3.1 with (2′) of the following type.

Algorithm 3.2. (1) Solve the nonlinear problem in coarse grid. This produces u_H, p_H and q_H.

(2) Take one block Gauss–Seidel step for the first two equations in fine grid using (u_H, p_H, q_H) as the starting point. This yields to $\bar{u}_h$, $\bar{p}_h$.

(3) Perform one Newton iteration for the whole system in fine grid using $(\bar{u}_h, \bar{p}_h, q_H)$ as the initial value. Get the solution (u_h, p_h, q_h).

This procedure gives an approximation nearly of order $\mathrm{O}(h^2)$ if the grid sizes $H = \sqrt{h}$ are chosen. For more details and examples, we refer to [25, 26].

References

[1] R. Acar, Identification of the coefficient in Elliptic Equations, *SIAM J. Control Optim.* **31** (1993) 1221–1244.

[2] G. Alessandrini, An identification problem for an elliptic equation in two variables, *Ann. Math. Pura Appl.* **145** (1986) 265–296.

[3] H.T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, (Birkhäuser, Basel, 1989).

[4] H.T. Banks and S. Reich and I.G. Rosen, An Approximation Theory for the Identification of nonlinear distributed parameter systems, *SIAM J. Control Optim.* **28** (1990) 552–569.

[5] V. Barbu, *Analysis and Control of Nonlinear Infinite Dimensional Systems* (Academic Press, New York, 1993).

[6] E. Casas, Optimal control in coefficients of elliptic equations with state constraints, *Appl. Math. Optim.* **1** (1992) 21–37.

[7] P.G. Ciarlet and J.L. Lions, Eds., *Handbook of Numerical Analysis, Vol 2: Finite Element Method* (North-Holland, Amsterdam, 1991).

[8] M. Crouzeix and V. Thomée, The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces, *Math. Comput.* **48** (1987) 521–532.

[9] R.S. Falk, Approximation of a class of optimal control problems with order of convergence estimates, *J. Math. Anal. Appl.* **44** (1973) 28–47.

[10] R.S. Falk, Error estimates for the numerical identification of a variable coefficient, *Math. Comput.* **40** (1983) 537–546.

[11] J. Frehse and R. Rannacher, Eine L^1- fehlerabschätzung für diskrete grundlösungen in der methode der finiten elementen, *Bonner Math. Scriften* **89** (1976) 92–114.

[12] T. Geveci, On the approximation of the solution of an optimal control problem governed by an elliptic equation, *RAIRO, Numer. Anal.* **13** (1979) 313–328.

[13] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, Berlin, 1983).

[14] J. Haslinger and P. Neittaanmäki, *Finite Element Approximations for Optimal Shape Design, Theory and Applications* (Wiley, New York, 1988).

[15] T. Kärkkäinen, *Error estimates for distributed parameter identification problems*, Univ. of Jyväskylä, Ph.D. dissertation, Department of Mathematics, Report 65, 1995.

[16] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and their Applications* (Academic Press, New York, 1980).

[17] R.V. Kohn and B.D. Lowe, A variational method for parameter identification, *RAIRO, Math. Model. Numer. Anal.* **22** (1988) 119–158.

[18] A.I. Koselev, A priori estimates in L^p and generalized solutions of elliptic equations and systems, *Engl. Transl. Amer. Math. Soc.* **20** (1962) 105–171.

[19] K. Kunisch and L.W. White, Parameter estimation, regularity and the penalty method for a class of two point boundary value problems, *SIAM J. Control Optim.* **25** (1987) 100–120.

[20] J.L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations* (Springer, Berlin, 1971).

[21] M.M. Mäkelä and P. Neittaanmäki, *Nonsmooth Optimization: Analyses and Algorithms with Applications to Optimal Control* (World Scientific Publishing Co., Singapore, 1992).

[22] K. Malanowski, Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems, *Appl. Math. Optim.* **8** (1981) 69–95.

[23] P. Neittaanmäki, M. Rudnicki and A. Savini, *Inverse Problems and Optimal Design in Electricity and Magnetism* (Oxford University Press, Oxford, 1995).

[24] P. Neittaanmäki and D. Tiba, *Optimal Control of Nonlinear Parabolic Systems; Theory, Algorithms and Applications* (Marcel Dekker, New York, 1994).

[25] A. Niemistö, L^p and $W^{1,p}$-error estimates for a class of inverse problems and their solving by two-grid methods, Univ. of Jyväskylä, Department of Mathematics, Laboratory of Scientific Computing, Report 9, 1995.

[26] A. Niemistö, L^p and $W^{1,p}$-error estimates for a class of saddle point problems, Univ. of Jyväskylä, Department of Mathematics, Laboratory of Scientific Computing, Report 8, 1995.

[27] R. Rannacher, On the convergence of the Newton–Raphson method for strongly nonlinear finite element method, in: *Nonlinear Computational Mechanics* (Springer, Berlin, 1991) 11–30.

[28] R. Rannacher and R. Scott, Some optimal error estimates for piecewise linear finite element approximations, *Math. Comput.* **38** (1982) 437–445.

[29] R. Scott, Optimal L^∞-estimates for the finite element method on irregular meshes, *Math. Comput.* **30** (1976) 681–697.

[30] G.M. Vainikko, Galerkin's perturbation method and the general theory of approximate methods for non-linear equations, *USSR Comp. Math. Phys.* **7** (1967) 1–41.

[31] K. Witsch, Projektive Newton-verfahren und anwendungen auf nichtlineare randwertaufgaben, *Numer. Math.* **31** (1978) 209–230.

[32] J. Xu, Two-grid finite element discretization for nonlinear elliptic equations, *SIAM J. Numer. Anal.*, submitted.